



Reflexive intensification in Spanish: Toward a complex reflexive?

Pedersen, Johan

Published in:
Historical Linguistics 2003, Selected Papers from the 16th International Conference on Historical Linguistics

Publication date:
2005

Document version
Peer reviewed version

Citation for published version (APA):
Pedersen, J. (2005). Reflexive intensification in Spanish: Toward a complex reflexive? In *Historical Linguistics 2003, Selected Papers from the 16th International Conference on Historical Linguistics* (Series IV ed., pp. 199-223). John Benjamins Publishing Company.

Unbound reflexive constructions in Spanish. A historical perspective¹

Johan Pedersen
University of Copenhagen

1. Introduction

This study aims to analyze the Spanish reflexive/intensifier distinction in a diachronic-typological perspective. Many languages have reflexive markers that can be combined with optional intensifiers.

This is exemplified in the following Spanish construction:

- (1) Pedro confía en sí (mismo)
Pedro relies on REFL self
“Pedro relies on himself”

This also applies to e.g. German, which is typologically similar to Spanish in this respect (Otero 1999: 1456):

- (2) Der Kanzler ist stolz auf sich (selbst) (König & Siemund 2000: 50)
The chancellor is proud of REFL self
“The chancellor is proud of himself”

It is generally assumed, thus, that in (1) and (2) the linguistic marker of reflexive meaning (*sí/sich*) has to be distinguished grammatically from the marker of intensification (*mismo/selbst*).² See e.g. König 2001, König & Siemund 2000, Otero 1999, Baker 1995, Kemmer 1995. In this paper, however, it will be argued, on the basis of empirical evidence, that the Spanish reflexive-intensifier construction *sí (mism-)* is in a process of being reanalyzed as a complex reflexive *sí mism-*.

Let me first point out what might be considered “the general view” on reflexive intensification in Spanish linguistics, as recently stated by Carlos Peregrín Otero in the extensive and detailed Spanish grammar: *Gramática Descriptiva del Español* (Otero 1999: 1449, 1509):

“The modifier *mism-* is always possible, but it is never necessary, ..., so we can deduce that it plays no role in the expression of reflexivity The distinction between morphologically simple and complex anaphors does not seem to have any consequences for Spanish that are worth mentioning” (my translation from Spanish).

¹ I owe thanks to the following persons for valuable comments and criticisms of previous drafts and presentations of this paper: Ulrich Detges, Lene Schøsler, Henning Andersen, Daniel Jacob, Jens Nørgård-Sørensen, Thora Vinther, Kasper Boye, Ronald Langacker, Nicole Delbeque, Margaret E. Winters, Concepción Company Company, Susana Silvia Fernández and to an anonymous reviewer of a shorter and introductory version of this paper that will be published as Pedersen (forthcoming).

² Henceforth I will refer to all possible variants of this expression by using the form *mism-*, e.g. *para sí mism-*. It should also be noted that, for reasons stated below, I have restricted the empirical part of this study to include only third person forms of the reflexive.

Basically he claims, maintaining the traditional point of view, that *mism-* has nothing to do with the expression of reflexivity in Spanish. See e.g. Alarcos Llorach 1995: 127; Real Academia Española 1981: 211; Franch y Blecua 1975: 677f; Bello & Cuervo 1954: 246. Defenders of this position are forced to face the following question: Is *mism-* always optional?

In Spanish reflexivity is basically expressed in two different constructions: the prototypical argument structure *se*-construction, in which *se* is a clitic reflexive, sometimes doubled by an unbound complex form in the prepositional case (*a sí mism-*), as in (3); and a non-clitic *sí*-construction, sometimes intensified by *mism-*, in which *sí* is an unbound reflexive in the prepositional case, as in (1):

- (3) Fernando **se** lava (a **sí mismo**)
 Fernando REFL washes OBJ MARK REFL self
 “Fernando washes himself”

- (1) Pedro confía en **sí** (mismo)
 Pedro relies on REFL self
 “Pedro relies on himself”

It is well known that in the prototypical argument structure construction, cf. (3), in which the clitic *se* is required, *mism-* is certainly not optional in Modern Spanish. You can’t say: **se lava a sí*; it has to be: *se lava a sí mism-*. If we look at the non-clitic constructions, in many contexts the complex form, including *mismo*, is required for most speakers (Otero 1999), as exemplified in (4). This is not the case in similar German constructions of intensification, in which the use of *selbst*, on the contrary, is restricted to the most extreme cases of non-other directed predicates (König & Siemund 2000: 63f; see also Otero 1999: 1456 and Baker 1995: 98f). This is exemplified in (5):

- (4) Pedro está orgulloso de **sí mismo**
 Pedro is proud of REFL self
 “Pedro is proud of himself”

- (5) Paul stolperte über sich selbst (König & Siemund 2000: note 18)
 Paul stumbled over REFL self
 “Paul stumbled over himself”

On the other hand, Danish, as well as other languages, has a complex reflexive with an incorporated (obligatory) element of intensification. In Danish you would always say:

- (6) Han sad og tænkte over det for **sig selv** (*han sad og tænkte over det for *sig*).
 He sat and thought about it for REFL self
 “He was thinking about it on his own”

Consequently, for the speakers concerned, Spanish *mism-* seems to some extent, just as the Danish *selv* in (6), to be relevant for the construction of reflexivity in certain contexts. If it is true that *mism-* is in a process of becoming an obligatory part of the prepositional construction of reflexivity in Spanish, we may expect that the use of complex constructions containing *mism-* has increased markedly as related to the use of simple constructions without *mism-*. In order to approve, or discard, the idea of a complex reflexive being under development in Spanish, I have carried out a quantitative diachronic study of the use of *mism-* in reflexive constructions. Thus, the principal empirical question I seek to answer in this paper is: Is there a clear indication of historical change in the use of *mism-* in reflexive constructions?

The paper is structured as follows: In Section 2 some major contributions to the analysis of reflexive intensification in the literature will be sketched out. The empirical method will be outlined in Section 3. Subsequently, results from the empirical study of the diachrony of *sí mism-* will be presented in Sections 4 and 5. In Section 6 the data will be analyzed and interpreted, basically by examining the plausibility of two distinct hypotheses. Finally, the results will be summarized and a conclusion will be drawn in Section 7.

2 *The construction of reflexivity and intensification*

In many European languages, including Spanish, the reflexive and the intensifier may be distinguished as regards form as well as syntactic distribution (e.g. German: *sich/selbst*, Italian: *sè/stesso*, Spanish: *sí/mism-*). In these languages the reflexive pronoun and the intensifier are combinable: *sich selbst*, *sè stesso*, *sí mism-* etc, cf. (7)-(10):

- (7) Der Kanzler **selbst** wird anwesend sein =Intensifier
The Chancellor self will present be
“The Chancellor himself will be present” (König & Siemund 2000: 50)
- (8) Der Kanzler ist stolz auf **sich** (selbst) =Reflexive (+ intensifier)
The Chancellor is proud of REFL self
“The Chancellor is proud of himself” (König & Siemund 2000: 50)
- (9) Él **mismo** dijo que no lo iba a hacer. =Intensifier
He self said that not it would do
“He himself said that he wasn’t going to do it”

- (10) *Pedro* está orgulloso de *sí* (mismo) =Reflexive (+ intensifier)
 Pedro is proud of REFL self
 “He is proud of himself”

Correspondingly, the term *emphatic reflexives* (*emphatics*), which is used in part of the literature (e.g. Kemmer 1995; Quirk *et al.* 1985), indicates that intensifiers are supposed to be understood merely as providing a special emphatic use of reflexives. Nevertheless, in theoretical terms, in particular when semantic aspects are to be taken into account, various scholars have argued that it is in some cases highly problematic to draw a clear distinction between reflexives and reflexive intensifiers (e.g. König & Siemund 2000, Baker 1995, Zribi-Hertz 1995). On the other hand, some recent theoretical studies of reflexivity distinguish between at least two types of reflexives: 1) morphologically simple (*simplex* or *SE-*) anaphors (Dutch *zich*, Danish *sig*, Italian *sè*, etc.) and 2) morphologically complex (*complex* or *self-*) anaphors (e.g. English *himself*, Dutch *zichzelf*, Danish *sig selv*) (see e.g. Burzio 1996, 1991; Reinhard and Reuland 1993; Pica 1991; Everaert 1986). In the latter type, the simple reflexive is typically obligatorily combined with an element of intensification, constituting together a complex reflexive³. Basically there are:

A) Simple reflexives

- (11) Hans zag de hond naast **zich** / ***zichzelf** Dutch (Everaert 1986: 2)
 Hans saw the dog next SIMP REFL / SELF-REFL
 “Hans saw the dog next to him”

B) Complex reflexives

- (12) Han er stolt af **sig selv** / ***sig** Danish
 He is proud of SELF-REFL / *SIMP-REFL
 “He is proud of himself”

C) Simple reflexives + optional intensifier

- (13) Der Kanzler ist stolz auf **sich** (selbst) German
 The chancellor is proud of REFL self
 “The chancellor is proud of himself”

It is rather unclear how intensifiers have developed historically. On the other hand, the typical case seems to be that reflexive anaphors develop from intensifiers (see e.g. König & Siemund 2000,

³ Notice, however, that English does not distinguish morphologically between reflexive and intensifier, e.g. *he washed himself* (reflexive) and *Peter himself*... (intensifier).

Schladt 2000, Baker 1995, Zribi-Hertz 1995). However, this typological pattern is not necessarily applicable to Modern Spanish. It is primarily based on the fact that in a number of languages (e.g. English, but not Spanish) the same expression is used both as an intensifier and as a reflexive anaphor. Moreover, in the Spanish case we have to take into account the fact that the prototypical reflexive construction has undergone a process of cliticization, as will be shown in Section 2.2.

2.1 *The meaning of reflexivity and intensification*

The intensifier identifies an entity that is salient in the discourse in contrast to other potential referents, which have just been mentioned, or could have been mentioned (Kemmer 1995: 66). König & Siemund claim, in line with Kemmer, that intensifiers provide potential alternatives to the accentuated NP-referent, to which they are typically collocated. The use of the intensifier implies that the status of the NP-referent has to be evaluated by inferring an actual prominence structure in discourse (König & Siemund 2000: 45-46, 1996a,b). See also Baker (1995: 80ff.) and König (1991: 87ff.). It seems clear that the construction of intensification is closely related to the construction of reflexive meaning, both diachronically and synchronically. Both constructions imply, in many cases, that something is contrary to the expectations created in discourse (Kemmer 1995). The reflexive marker typically encodes a coreference that is contrary to the expectation created in the semantic valency structure of the verb, or in context (*The old horse heaved himself out of the mud*, Kemmer 1995: 56). The intensifier, in turn, often appears emphasizing, unexpectedly, the prominent role in the clause, or in the discourse, of some referent as opposed to more natural candidates (*I wanted Marie herself to tell me*, Kemmer 1995: 56). This common semantic structure of reflexivization and intensification highlights the conceptual closeness of the categories intensifier and reflexive marker. The term “unexpectedness”, however, seems to be problematic, since the intensifier is not basically used, according to Baker, to indicate unexpectedness but to mark discourse prominence (Baker 1995: 95, note 33). See also the discussion in Zribi-Hertz 1995 and König & Siemund 2000.

2.2 The clitic argument structure construction in Spanish

In modern Spanish the unbound reflexive, as direct or indirect object (e.g. *a sí mism-*), has to be doubled by a clitic reflexive (*se*). In other words, the use of the clitic is obligatory in these constructions (for more details, see Torrego 1995):⁴

- (14) **Se** ha tomado el pulso (a **sí mismo**) Modern Spanish
REFL has taken the pulse OBJ MARK REFL self
“He has examined his own pulse”

On the other hand, this use of the unbound reflexive is only possible in some verbal constructions, and it is never obligatory (in brackets in (14)). Therefore, we may hypothesize in the first place that in those reflexive constructions in which the unbound form may occur as a redundant object marker, the whole expression ‘*a + REFL + mism-*’, not only *mism-*, has to be interpreted as an optional marker of reflexive intensification.

Notice, in addition, that the unbound form in these argument structure constructions of Modern Spanish only very rarely occurs without *mism-*, cf. (15), as in other prepositional case constructions, e.g. *para sí (mism-)*. *A sí* without *mism-* occurs only in adverbial prepositional expressions like *frente a sí*, in which there is no reflexive object to be doubled, cf. (16):

- (15) **Se** ha tomado el pulso *a **sí**

- (16) Clinton tiene frente a **sí** 4 años para innovar... (1996) (CREA)
Clinton has front of REFL 4 years to innovate...
Clinton has in front of him 4 years to innovate...

It seems therefore plausible to further hypothesize that in Modern Spanish the use of *mism-*, as a compulsory part of the optional intensifier in the clitic construction (*se... (a sí mism-)*), somehow is related to the selfsame cliticization of the construction.

In fact, in Old Spanish non-clitic constructions it was quite normal that the reflexive object was constructed as *a sí*, without *mism-*, though the use of *a sí mism-* was completely normal as well:

- (17) E obligaron a **sí** e a sus bienes, ...
and mortgaged-3PL OBJ MARK REFL and OBJ MARK their belongings
“and they mortgaged themselves and their belongings...” (CORDE, from 1308)

⁴ It is a general rule that unbound personal pronouns with object function have to be doubled by a clitic pronoun.

- (18) , y que el que ama **a** **sí mismo**, ... (CORDE, from 1537)
 and that DET that loves OBJ MARK REFL self
 “..., and that those who love themselves”

It follows that the increasing use of *mism-* in Old Spanish cannot only be due to the cliticization of the construction at that time. It might also be related to changes in the use of the prepositional object marker *a*. In Old Spanish one of the basic functions of the *a*-marker was to focalize human referents considered particularly important (Detges 2001). According to Detges, referring to Meier (1945: 247) and Delille (1970: 34), the use of the equivalents of the Spanish reflexive emphatic *mism-* in Medieval Portuguese and Old Catalan was normally an impediment for the use of the prepositional *a*-marker. For instance, when preceding *si mateix* in Catalan the preposition *a* was never used. Detges interprets this pattern as indicating that the focalizing use of the prepositional marker was considered unnecessary when other focalizing elements were used. Thus, we may similarly hypothesize that the loss of the focalizing weight of the Spanish *a*-marker in Old Spanish motivated an increased use of the emphatic element *mism-*, which, as we have seen, has become a compulsory part of the intensifying reflexive expression *a sí mism-* in the argument structure construction in Modern Spanish. This interpretation is actually highly compatible with, and supports, the hypothesis stated earlier in this section; the increased use of *mism-* in clitic constructions (*se...a sí mism-*) is related to the cliticization of the construction. This is so, because the focalizing weight of the *a*-marker was in fact weakened by the emergence of the clitic duplication itself (see the discussion in Detges 2001).

To summarize: In Modern Spanish the complexity (with *mism-*) of the reflexive unbound form is compulsory, whereas its use in the clitic argument structure construction is optional: *se (a sí mism-)*. Correspondingly, *mism-* seems to have achieved a new status in this construction in Modern Spanish: from being *mism-* an optional part of the reflexive marker in Old Spanish, the whole expression *a sí mism-* has in Modern Spanish an intensifying function in the clitic construction. In Section 6, data that support this interpretation will be presented. Another interesting perspective of this analysis is that both similar pragmatic factors and the syntactic pattern itself (the required complexity of the unbound reflexive form *sí mism-*) might have been motivating factors for an increased use of the

complex form *sí mism-* in non-prototypical prepositional constructions (cf. next Section). On the basis of extensive empirical evidence, this question will be analyzed in detail in Section 6.

2.3 The Spanish non-clitic reflexive

In the non-clitic construction the intensifier (*mism-*) is, according to Spanish convention, optional:

- (4) Pedro está orgulloso de *sí* (mismo)
Pedro is proud of REFL self
“Pedro is proud of himself”

However, many Spanish speakers prefer the complex reflexive form in many contexts. Carlos Otero (1999: 1500) mentions in passing that the complex form (+*mism-*) is required for some speakers in specific construction types, such as the one in (19):

- (19) Sólo Ana habla siempre de *sí* misma
Only Ana talks always about REFL self
“Only Ana is always talking about herself”

When the same speakers choose a non-reflexive anaphoric construction in the same context the complex form is not required:

- (20) Sólo Ana habla siempre de *ella* (misma)
Only Ana talks always about PERS PRON-3SG self

This is surprising since the pronominal reference in (20) actually is ambiguous in the simple anaphoric construction without *misma* (the pronominal anaphor refers either to Ana or to someone else), while it would not be so in a simple reflexive construction in (19). In other words, *misma* would resolve the ambiguous interpretation in (20), whereas there is no ambiguity to resolve in (19). Such “test” constructions suggest that *mism-* is not merely an intensifier, but is rather closely related to the construction of reflexive meaning.⁵ In the subsequent sections we will present substantial empirical evidence that shows that an increased use of the complex reflexive should not be ascribed simply to an increased, pragmatically determined, use of the intensifier. We will argue that such a change has to be seen as related to the development of a complex reflexive.

⁵ Note that *sólo* in (19) and (20) creates contextual conditions that make the realization of the *mism*-variant more compelling. Nevertheless, this is so in both constructions, and the point here is that if we compare the preference of the speakers, with respect to the use of *mism-*, in these two construction types, for some speakers the reflexive construction seems to be particularly well suited for the use of the complex form.

3 The empirical method: varying token ratio

In the empirical part of this study we have measured the use of the complex form *sí mism-* as related to the use of the simple form *sí*.; that is, the token ratio (TR) of the complex form. The token ratio has been calculated on the basis of searches in the diachronic corpus CORDE (C1) and the synchronic corpus CREA (C2) for uses of *sí* and *sí mism-*. Thus, needless to say, conclusions may be drawn only for the usage in these specific corpora. Nevertheless, if we consider the size and the composition of the corpora CORDE and CREA,⁶ available in detail on the website, it becomes clear that they have been composed with the objective of being representative of Old Spanish and Modern Spanish respectively. The basic unit of measurement (TR) is calculated as the fractional relationship between the occurrences of the two forms in a corpus C:⁷

$$\text{TR}(C) = \text{occurrences of 'preposition + } \textit{sí mism-} \text{' in } C / \text{occurrences of 'preposition + } \textit{sí} \text{' in } C$$

If $\text{TR}(C) > 1$ this is an indication of relatively more occurrences of the complex construction than the simple construction in corpus C. The difference between the token ratio of *sí mism-* in two corpora C1 and C2 is calculated as a differential factor DTR (difference of token ratio):

$$\text{DTR}(C1, C2) = \text{TR}(C2) / \text{TR}(C1)$$

If $\text{DTR}(C2, C1) > 1$ this is, *ceteris paribus*, an indication that there is a more extended use of the complex *sí mism-* in C2 than in C1. Let us say that C1, C2 are CORDE (diachronic) and CREA (Modern Spanish) respectively; then $\text{DTR}(C2, C1) > 1$ would indicate an increased use of the complex form in Modern Spanish as compared to Older Spanish.

The advantage of having the token ratio, instead of absolute values, as the basic measure of frequency and historical change, is the possibility of comparing non-homogeneous corpora of different sizes as a reliable source for the extraction of data. Dialectal and text-stylistic variation may subsequently be taken into account by analyzing usage in differentiated corpora. The problem is that no-

⁶ Each corpus contains approximately 120 million words. In CORDE the texts are from the beginning of the Spanish language to 1974. CREA contains Modern Spanish texts from 1975-.

⁷ In some cases substantial changes of very small values of TR may be poorly represented by the use of graphs or columns. In such cases the inverse token ratio, that is: 'preposition + *sí* / 'preposition + *sí mism-*', will be used in order to improve the visualization of the change, since the inverse formula will convert small values into corresponding high values.

table tendencies observed in the data may be due to over-representation of certain text types, or dialectal variants, in which the simple or the complex form is more or less frequent than elsewhere.

Another issue of corpus linguistic methodology that has to be taken into account is that this study has only included the third person forms of the reflexive construction.⁸ That is, we have only searched for occurrences of *sí mism-*, while *mí mism-* (first person) and *ti mism-* (second person) have been excluded from the study. We have made this choice for practical reasons. Since CORDE and CREA are not tagged corpora we have to search for words, not constructions. Only the third person forms (*sí*) are unambiguously reflexive: *compró la tarta para sí (mismo)* [*he bought the cake for REFL (self)*]. Searches for first and second person forms, e.g. *mí* or *ti*, would also include non-reflexive constructions: e.g. not only *lo hago por mí (mismo)* [*I do it for REFL (self)*], but also *no te preocupes por mí* [*don't worry about me = PERS PRON-1SG-PREP CASE*]; and not only *¿compraste la tarta para ti (mismo)?* [*did you buy the cake for REFL (self)*], but also *¿Pedro compró la tarta para ti?* [*did Pedro buy the cake for you = PRON-2SG-PREP CASE*]. There is no doubt that by excluding first person and second person forms the study is designed primarily for written language data, since spoken language to a larger extent than written language involve first and second person forms. Notice, however, that in languages that have reflexive pronouns in all three persons, as in Spanish, the use of reflexive constructions in the third person is much more frequent than in the first or second (Newmeyer 2003: 695). Moreover, shortcomings of limiting the study to third person forms are compensated for by using the token ratio as the basic measure of frequency. This method makes it possible to compare the use of the complex form in spoken and written language even so, provided that the basic corpus, from which the data are extracted, is big enough.

The chi square test (χ^2) has been used to ensure that observed variations really are significant, and that they should not be interpreted simply as incidences of statistical coincidences. As a statistical zero-hypothesis, the token ratio of *sí mism-* is taken to be constant, which is compatible with the

⁸ Methodological shortcomings imply that occurrences of fixed expressions, such as *de por sí*, are registered in data as instances of the simple form. As they do not allow for optional intensification (**de por sí mism-*) they are irrelevant for this study and should in ideal conditions have been left out.

general assumption in Spanish linguistics (see the introduction). According to the alternative hypothesis the token ratio is not constant, it is changing. The expected value of the zero-hypothesis is calculated on the basis of all the occurrences in CORDE and CREA, and the alternative values are calculated on the basis of the observed occurrences in subcorpora of limited ranges of time. The outcome of the calculation is a measure of the probability of making a mistake when postulating that the token ratio is different from the expected value predicted by the zero-hypothesis.

3.1 Formal variations of *sí mism-*

Mism- originates from Vulgar Latin *medīpsīmus*, a combination of *īpsīmus* (emphatic form of *ipse* = *self*) and *-med*, a colloquial variant of *-met*, which was added to the personal pronouns in order to strengthen their meaning (e.g. *egomet*, *tumet*); see e.g. Otero (1999), Fernández Ramírez (1987), Corominas (1954-57, under *mismo*). The form *mism-* was used in the *Poema de Mio Cid* (early 13th century) and in other early texts of Old Spanish. However, even though some of these uses of *mism-* might have been authentic, most of them were modernizations by the scribes (Corominas 1954-57). In this early period there are different formal variants, in particular *meísmo*, *misme* and *me(e)smo*. According to Corominas, the most common form in the Middle Ages, especially from the 14th century, was *mesmo*. From the beginning of the 17th century *mismo* is the dominant form, and from the 18th century *mesmo* is only sporadically used in rural areas. Searches in CORDE confirm that *mesmo* was used in the Middle Ages since this form occurs to some extent in CORDE, but almost exclusively from 1400-1700, as stated by Corominas. The other formal variants, on the other hand, are practically non-existent in CORDE (there is no, or only one, occurrence). The conclusion is that we need to adjust the data for occurrences of *mesm-*.

4 General data on the historical change

In Table 1 the general data on the historical change are presented for the most common prepositions used in this construction type: *a sí mism-* [*OBJ MARK REFL self*], *para sí mism-* [*to/for REFL self*], *por sí mism-* [*by REFL self*], *de sí mism-* [*of REFL self*], *en sí mism-* [*in REFL self*], *sobre sí mism-*

[*about REFL self*] and *consigo mism-* [*with REFL self*]. In the last table row the values are calculated on the basis of added occurrences of ‘preposition + *sí*’ and ‘preposition + *sí mism-*’ respectively, in order to provide a measure of the general tendency. The χ^2 -test confirms that there are actually significant historical changes going on for all the prepositional constructions. The data show that the token ratio (TR) has increased dramatically for all the prepositional constructions, though the change in using *a sí mism-* in the clitic construction is particularly striking. Also *sobre sí mism-* and *de sí mism-* have turned out to be predominant in the non-clitic constructions, which is reflected in the calculation of TR to be nearly 4,5 and 1,7 respectively in the Modern Spanish corpus (CREA). From the added occurrences the general value of TR in CREA has been calculated to be 1,38, which indicates that the complex form *sí mism-* is substantially more common than the simple form *sí* in the corpus of Modern Spanish. This is, on the contrary, definitely not so in the diachronic corpus (CORDE), for which TR is calculated to be 0,34. Finally, we should also pay attention to the differential factor DTR. *A sí mism-* has a DTR of more than 18, and as for *de sí mism-*, for example, the TR has been multiplied by more than four in the modern corpus, which is also the average factor for the added occurrences.

	Corde		TR	χ^2	Crea		TR	χ^2	DTR
	Sí	Sí mism-			Sí	Sí mism-			
A= obj marker	2281	3388	1,49	7,8E-197	173	4643	26,84	2,5E-231	18,07
Para = to/for	2522	261	0,10	2,41E-22	1166	450	0,39	2,77E-37	3,73
Por = by	6061	2264	0,37	2,79E-42	2465	2184	0,89	2,75E-74	2,37
De = of	7009	2951	0,42	3,9E-107	1754	2999	1,71	2,7E-222	4,06
En = in	6230	1912	0,31	2,88E-98	2218	2552	1,15	2,5E-166	3,75
Sobre = about	967	444	0,46	0,000768	164	727	4,43	4,11E-05	9,65
Con = with	9930	690	0,07	2,83E-52	2534	873	0,34	6,1E-159	4,96
Added occurrences	35000	11910	0,34	0	10474	14428	1,38	0	4,05

Table 1 Token ratio of *sí mism-* in an overall historical perspective

Observed changes may be due to overrepresentation in a corpus of certain text types for which the simple, or the complex form, is more frequently used.⁹ This is why calculations of the token ratio in general terms have to be supplemented with calculations based on differentiated corpora.

4.1 Text type variation

The data might reflect a different use of *sí mism-* in different text types, due to overrepresentation of certain text types in which the simple or the compound form is more frequently used. It is therefore necessary to take into consideration the possibility that the use of *si mism-* varies in different text types and in spoken versus written language. There may be dialectal variation as well. If the composition of the corpora being compared is not the same, as regards the text types and dialects represented, such variation may influence the calculated TR for the complex *sí mism-* construction. One way to account for this variation is to examine the development of the TR for specific text types. In Table 2 I have compared the TR in corpora of spoken and written modern Spanish. Table 3-5 represent a differentiation made on the basis of comparable text categories in the corpora CORDE and CREA. Surprisingly, there is no indication at all that the complex form should be more frequent in spoken language than in written language. The distinction between written and spoken language seems, thus, to be a rather marginal factor, and there is no way it could reasonably be ascribed any substantial explanatory role in the present study. We will come back to this issue below.

	Spoken language	Written language
A	34,00	26,71
Para	0,65	0,43
Por	0,54	0,91
De	0,96	1,73
En	0,39	1,26
Sobre	8,00	4,41
Con	0,20	0,35

Table 2 TR of preposition + *sí (mism-)* in spoken/written Modern Spanish.

⁹ This may also be true for the distinction of written vs. spoken language; see below.

By comparing similar categories of newspaper language, fictional language and non-fictional language in CORDE and CREA we have found that for almost every single text type the use of the complex form has increased drastically. See Table 3-5:

	CORDE		CREA
	Newspaper	Newspaper prose	Newspaper
A	0,81	0,83	50,95
Para	0,03	0,06	0,42
Por	0,21	0,26	0,69
De	0,26	0,30	1,66
En	0,28	0,24	1,33
Sobre	0,12	0,19	3,79
Con	0,04	0,02	0,27

Table 3 TR of preposition + *sí* (*mism-*) in newspaper-Spanish. A historical perspective.

	CORDE		CREA
	Lyrical prose	Drama	Fiction
A	1,69	3,19	19,33
Para	0,13	0,16	0,38
Por	0,35	0,43	0,96
De	0,49	0,78	1,52
En	0,20	0,58	1,13
Sobre	0,71	0,87	4,14
Con	0,08	0,23	0,49

Table 4 TR of preposition + *a sí* (*mism-*) in Spanish fictional prose. A historical perspective.

	CORDE			CREA		
	Society	Didactics	Science	Science / society	Trade / Politics / Finance	Science / technology
A	1,21	1,89	1,83	27,13	22,21	31,50
Para	0,12	0,11	0,05	0,35	0,17	0,47
Por	0,53	0,55	0,33	1,36	0,69	0,8
De	0,33	0,50	0,30	2,48	1,42	1,25
En	0,29	0,44	0,42	1,42	1,48	0,92
Sobre	0,13	0,97	0,42	4,60	2,29	–
Con	0,07	0,09	0,05	0,38	0,15	0,05

Table 5 TR of preposition + *sí* (*mism-*) in Spanish non-fictional prose. A historical perspective.

In general, we found that the use of *sí mism-* has increased markedly in all text types, though some variation can be observed, as is only to be expected. As far as the dialectal variation is concerned,

the tendency of increased use of the complex form is clearly strongest in Iberian Spanish. Nevertheless, such variation has not been examined systematically.

5 *More specific data on the historical change*

In this section, we will show in more detail how the use of the complex construction ‘preposition + *sí mism-*’ has developed. As illustrated in Table 6 and Figure 1-2, it has been undergoing constant change up to present time, though the change from the fifteenth to the sixteenth century is particularly remarkable. The corpus (CORDE) has very few occurrences in the period 0-1400, in particular of the complex form; in some cases (with specific prepositions) only a few, or even no occurrences at all. Therefore, we have to be particularly careful when we interpret the data from this period. This is why the discussions and interpretations of the data presented in the subsequent sections are based exclusively on data from 1400-.

Sí (mism-)	Si	Sí mism-	TR	Inverse values	DTR
0–1300	669	37	0,055	18,08	
1300–1400	592	20	0,03	29,60	
1400–1500	2752	94	0,03	29,28	
1500–1600	11278	1451	0,13	7,77	
1600–1700	4484	902	0,20	4,97	
1700–1800	1121	283	0,25	3,96	
1800–1900	3800	1819	0,48	2,09	
1900–1975	7105	6994	0,98	1,02	
1975- (CREA)	10474	14428	1,38	0,73	4,05
CORDE (total)	35000	11910	0,34	2,94	

Table 6 The development of *sí mism-* in terms of token ratio

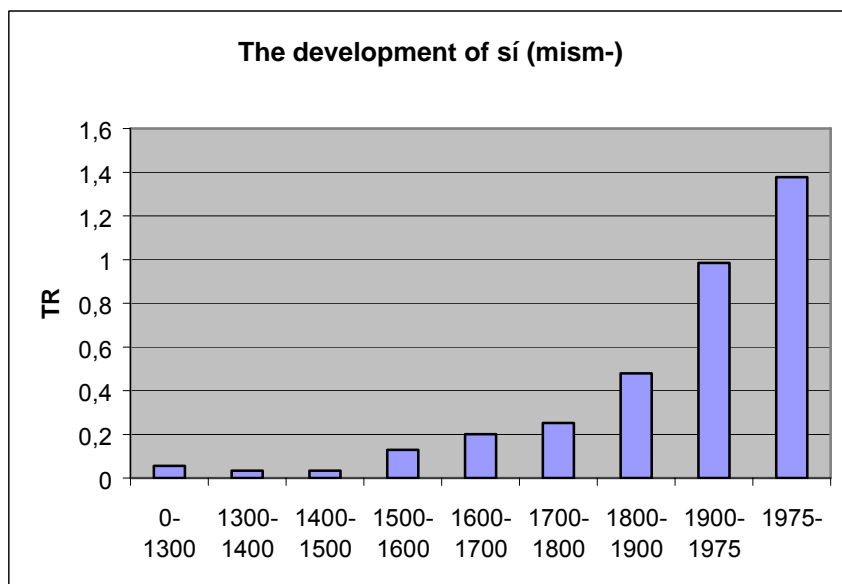


Figure 1 The development of *sí mism-* in terms of token ratio

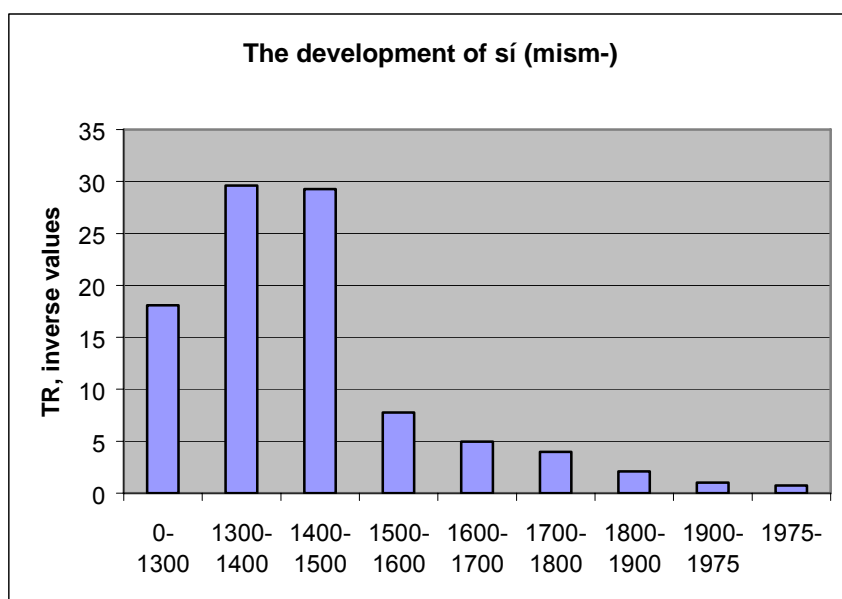


Figure 2 The development of *sí mism-* in terms of inverse token ratio.

Subsequently, the data for each preposition will be presented.

A	Si	Sí mism-	TR(C)	χ^2
0–1300	35	29	0,83	3,41E-09
1300–1400	27	7	0,26	1,23E-14
1400–1500	283	33	0,12	9E-170
1500–1600	896	434	0,48	0
1600–1700	383	320	0,84	2,39E-84
1700–1800	53	94	1,77	0,000292
1800–1900	180	415	2,31	7,99E-05
1900–1975	286	1999	6,99	9,88E-35
1975–	173	4643	26,84	2,5E-231

Table 7 The development of *a sí mism-* in terms of token ratio.

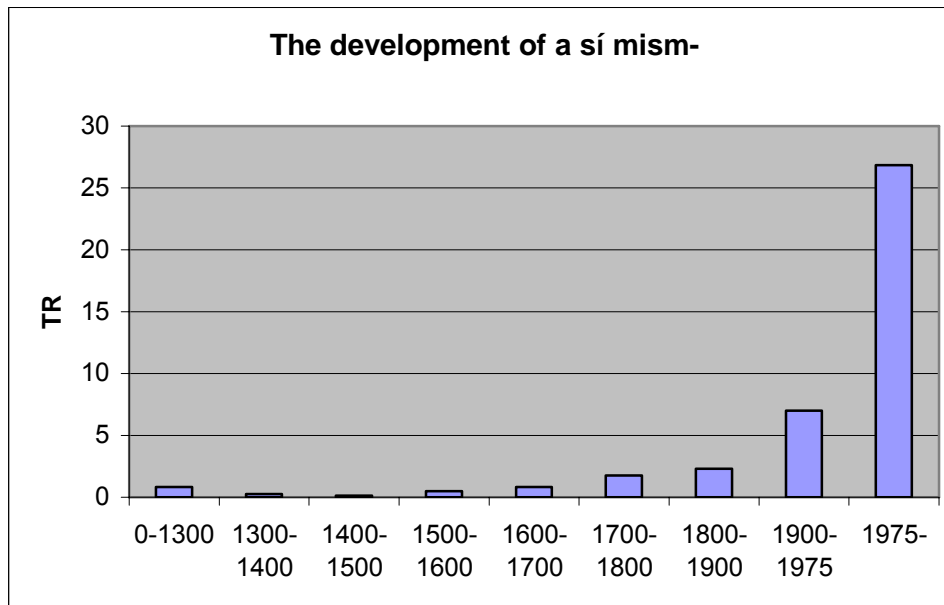


Figure 3 The development of *a sí mism-* in terms of token ratio.

Para	Si	Sí mism-	TR(C)	χ^2
0–1300	14	0	0	0,10041
1300–1400	21	2	0,10	0,330634
1400–1500	153	1	0,01	1,7E-07
1500–1600	737	33	0,04	3,45E-19
1600–1700	371	22	0,06	1,27E-08
1700–1800	90	6	0,07	0,008328
1800–1900	334	15	0,04	1,73E-09
1900–1975	596	176	0,30	5,49E-07
1975–	1166	450	0,39	2,77E-37

Table 8 The development of *para sí mism-* in terms of token ratio.

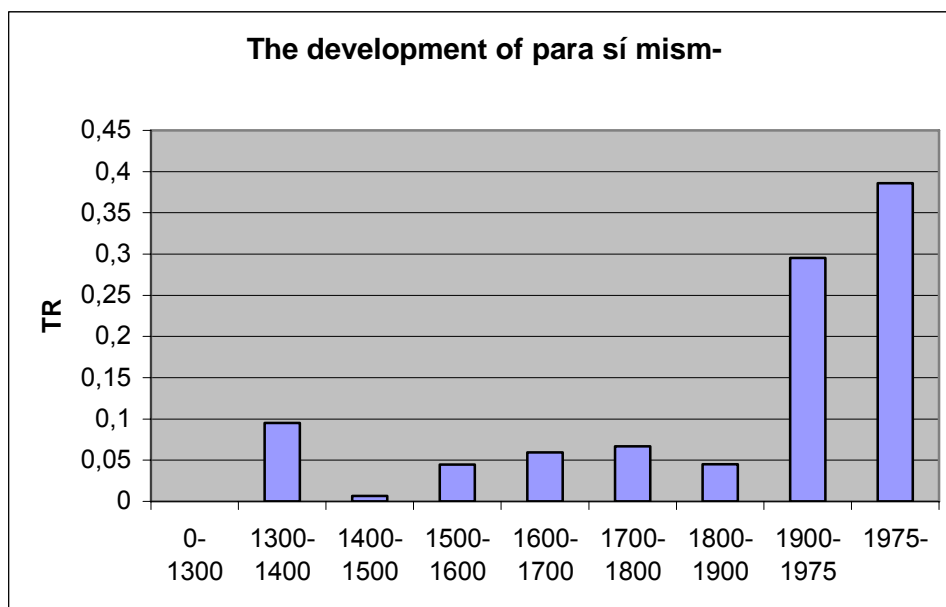


Figure 4 The development of *para sí mism-* in terms of token ratio.

Por	Si	Sí mism-	TR(C)	χ^2
0-1300	165	1	0,01	6,1E-20
1300-1400	113	4	0,04	2,01E-12
1400-1500	332	16	0,05	1,88E-31
1500-1600	1827	201	0,11	2,7E-118
1600-1700	707	101	0,14	6,74E-39
1700-1800	250	73	0,29	9,7E-06
1800-1900	976	506	0,52	0,909018
1900-1975	1489	1229	0,83	3,21E-33
1975-	2465	2184	0,89	2,75E-74

Table 9 The development of *por sí mism-* in terms of token ratio.

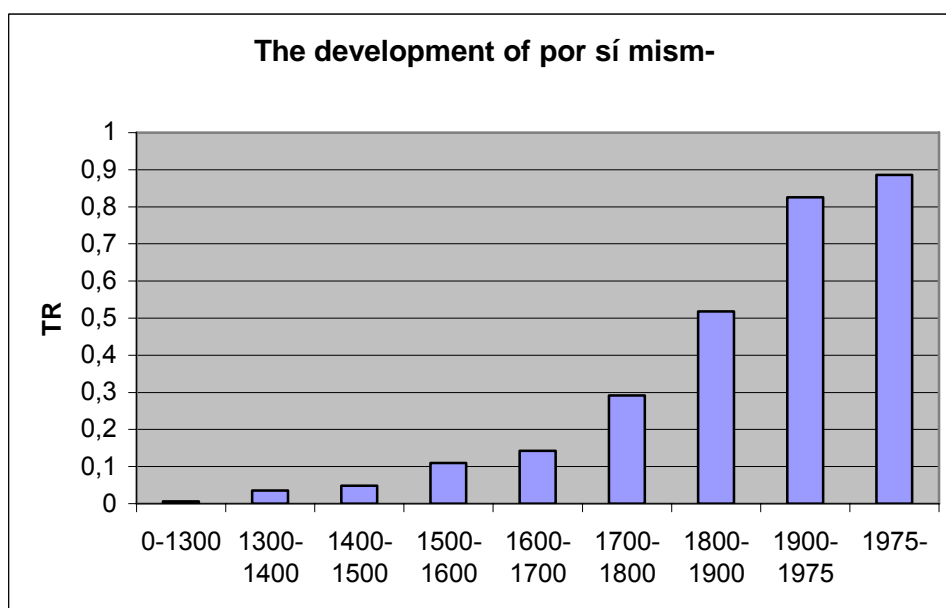


Figure 5 The development of *por sí mism-* in terms of token ratio.

De	Si	Sí mism-	TR(C)	χ^2
0-1300	47	3	0,06	6,97E-07
1300-1400	53	5	0,09	7,9E-07
1400-1500	392	16	0,04	4,66E-51
1500-1600	2528	489	0,19	5,6E-162
1600-1700	950	316	0,33	3,15E-29
1700-1800	220	63	0,29	4,63E-10
1800-1900	730	380	0,52	2,52E-05
1900-1975	1413	1627	1,15	7,05E-49
1975-	1754	2999	1,71	2,7E-222

Table 10 The development of *de sí mism-* in terms of token ratio.

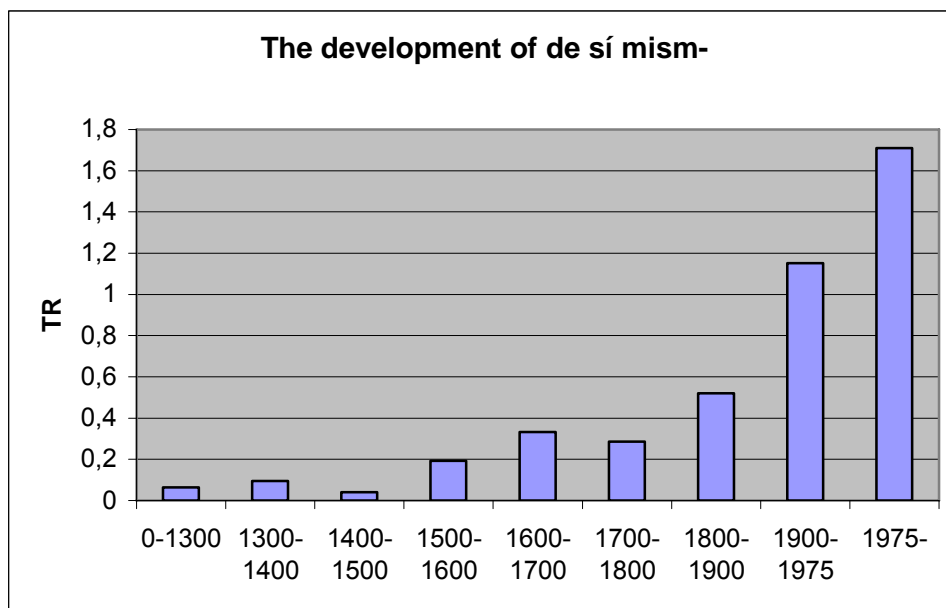


Figure 6 The development of *de sí mism-* in terms of token ratio.

En	Si	Sí mism-	TR(C)	χ^2
0-1300	22	3	0,14	0,017643
1300-1400	77	2	0,03	2,13E-09
1400-1500	344	21	0,06	5,42E-31
1500-1600	2194	208	0,09	4,3E-157
1600-1700	797	108	0,14	1,65E-46
1700-1800	188	29	0,15	5,07E-11
1800-1900	618	296	0,48	0,164398
1900-1975	1547	1201	0,78	7,86E-24
1975-	2218	2552	1,15	2,5E-166

Table 11 The development of *en sí mism-* in terms of token ratio.

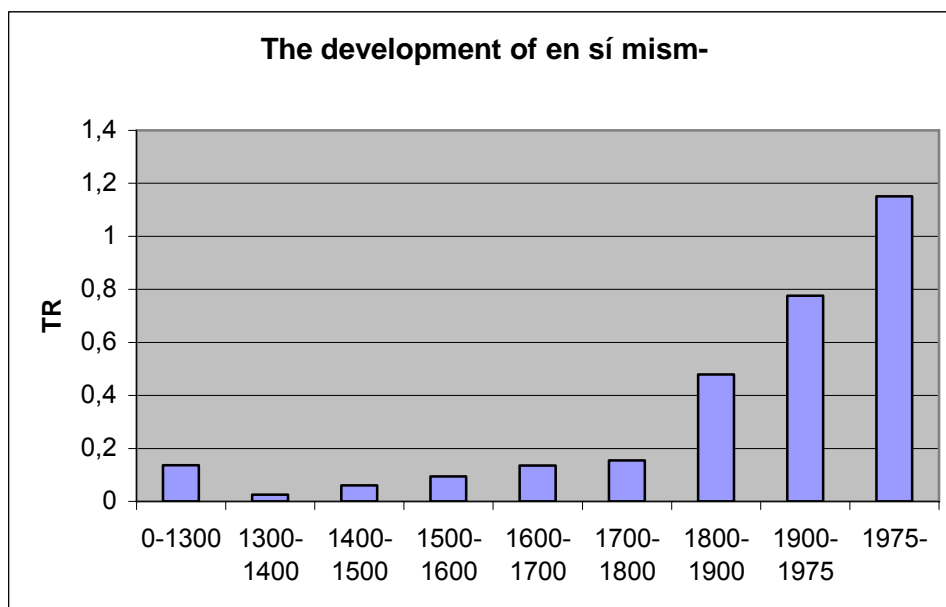


Figure 7 The development of *en sí mism-* in terms of token ratio.

Sobre	Si	Sí mism-	TR(C)	χ^2
0-1300	15	0	0	8,12E-05
1300-1400	5	0	0	0,02289
1400-1500	75	0	0	1,23E-18
1500-1600	298	8	0,03	5,83E-64
1600-1700	107	2	0,02	1,31E-24
1700-1800	55	6	0,11	1,45E-10
1800-1900	154	84	0,55	1,54E-06
1900-1975	235	333	1,42	0,000217
1975-	164	727	4,43	3,6E-75

Table 12 The development of *sobre sí mism-* in terms of token ratio.

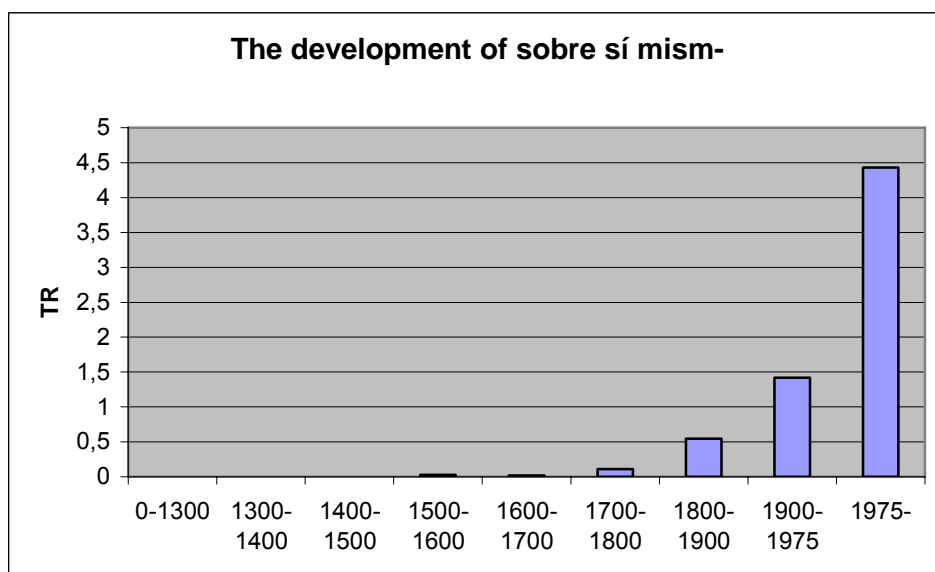


Figure 8 The development of *sobre sí mism-* in terms of token ratio.

Con	Sigo	sigo mism-	TR(C)	χ^2
0-1300	371	1	0,003	2,64E-11
1300-1400	296	0	0,002	1,11E-09
1400-1500	1173	7	0,01	1,09E-30
1500-1600	2798	78	0,03	8,15E-47
1600-1700	1169	33	0,03	2,2E-20
1700-1800	265	12	0,05	0,000315
1800-1900	808	123	0,15	0,044846
1900-1975	1539	429	0,28	5,17E-51
1975-	2534	873	0,34	6,1E-159

Table 13 The development of *consigo mism-* in terms of token ratio.

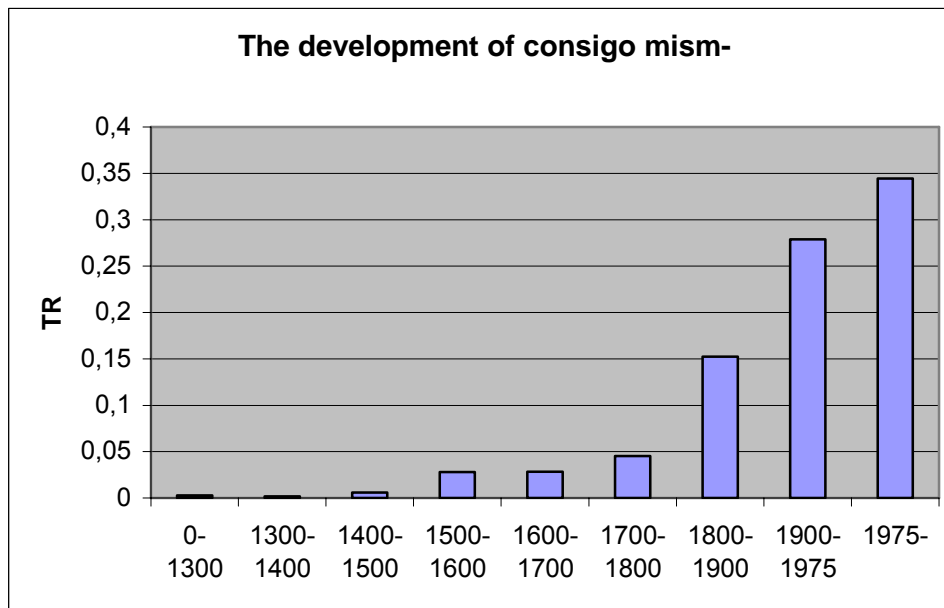


Figure 9 the development of *consigo mism-* in terms of token ratio.

There are several tendencies that we should pay special attention to. In the first place, for all prepositional constructions the corpora show a continuous tendency during the last 700 hundred years toward an increased use of the complex form. Regarding *a sí mism-*, the compound form is almost the only option in Modern Spanish. As for *de sí mism-* and *sobre sí mism-*, the complex forms have become dominant as well. For almost all the prepositional constructions, the process of change shows a dramatic change from the fifteen to the sixteen century; though this remarkable shift is in some of the cases poorly represented by the graphics; e.g. in the case of *a sí mism-*. Such representational problems can easily be solved, as mentioned earlier, by constructing the graphics on the basis of inverse values.

6 *Interpreting the data*

The data show a continuous, not abrupt, change toward an increased use of the complex form. This study supports, therefore, the claim that language change is a gradual process (see the discussion in Croft 2000: 49ff and references cited there). Gradualness is difficult to account for in child-based generative theories of language change, although some important attempts have been made (e.g. Lightfoot 1991). Among non-generative approaches a considerable number of frameworks are not based on child-acquisition, but rather on adult usage of language (see e.g. Milroy 1992, Bybee et al 1994, Keller 1994, Croft 2000). In such frameworks, often referred to as usage-based, the gradualness of linguistic change is certainly expected and handled straightforwardly. Moreover, changes in use are supposed in these frameworks to have frequency effects on grammar (Bybee et al 1994, Bybee & Hopper 2001). In that respect an increased use of the intensifier *mism-* would be expected to alter the status of this item in grammar. In this section we will be concerned with how to interpret the data by evaluating basically two types of usage-based analysis (cf. the discussion in Newmeyer 2003: 694; Croft 2000: chap. 5; Labov 1994: chap. 20):

1. *A paradigmatic interpretation* (Section 6.1) according to which the selection and the usefulness of an item is in focus as a decisive parameter for change. A paradigmatic interpretation of the data presented in this paper would be of the following type: The optional use of *mism-* as an intensifier may have increased gradually during the last 700 years for discourse-functional reasons.
2. *A syntagmatic interpretation* (Section 6.2), form-function reanalysis (Croft 2000: 120).

Recall that in processes of grammaticalization paradigmatic changes affect only one element in the grammaticalizing construction, whereas syntagmatic changes affect more than one element (cf. Lehmann 1985, Haiman 1985). In a syntagmatic perspective we will, in order to offer an explanation of the data, consider the possibility that *mism-* has been in a process of becoming gradually reanalyzed as a compulsory part of the reflexive marker.

6.1 A paradigmatic interpretation

First, we may hypothesize that the increase in token ratio of *sí mism-* reflects a general rise in the selection of *mism-* as a useful device for intensification. Such an interpretation of the data would predict that a similar phenomenon should be observable in constructions in which *mism-* is generally assumed to function as an intensifier. Another aspect of this hypothesis is that if the increased use of the complex form exclusively is related to the selection of a useful element, this inherently implies an increased speaker involvement. Therefore, if the hypothesis of usefulness is correct, it is predicted that the observed change should be notably distinct in spoken language, when compared to written language, since the speaker involvement in spoken discourse may be assumed to be relatively higher.

6.1.1 Distributional data

The first prediction is testable by examining how the token ratio of the complex form *él/ella mism-* [*he/she self*]¹⁰ has developed in the construction type: *él (mism-) lo ha visto* [*he (self) it has seen*] in which it is generally accepted that *mism-*, and equivalents in typologically similar languages likewise, functions as an intensifier (see e.g. König & Siemund 2000). By measuring the historical development of the expression *él/ella (mism-)*, we will get an indication of how the use of *mism-* as an intensifier of (pro)nominal phrases has changed, if it has. Thereby, we will be able to evaluate whether or not the altered use of the reflexive construction *sí (mism-)* could plausibly be ascribed to the general usefulness of *mism-* as an intensifier in communication. The results from this part of the study are tabulated in Table 14 and represented in Figure 10:

¹⁰ Reflecting the use of the complex form as related to the use of the simple form.

	El/ella	El/ella mism-	TR	DTR	χ^2
0-1975 (CORDE)	149233	2003	0,013		3,26E-07
1975- (CREA)	260366	1853	0,023	1,72	1,14E-15
0-1300	7916	74	0,009		1,21E-38
1300-1400	6643	27	0,004		9,54E-15
1400-1500	30363	215	0,007		7,50E-14
1500-1600	72733	950	0,013		1,06E-11
1600-1700	40989	493	0,012		1,31E-27
1700-1800	12426	309	0,025		1,50E-99
1800-1900	42237	1006	0,024		4,15E-62
1900-1975	91678	2389	0,026		3,2594E-07
1975- (CREA)	260366	5504	0,021		1,1407E-15

Table 14 The development of *él/ella mism-* in terms of token ratio

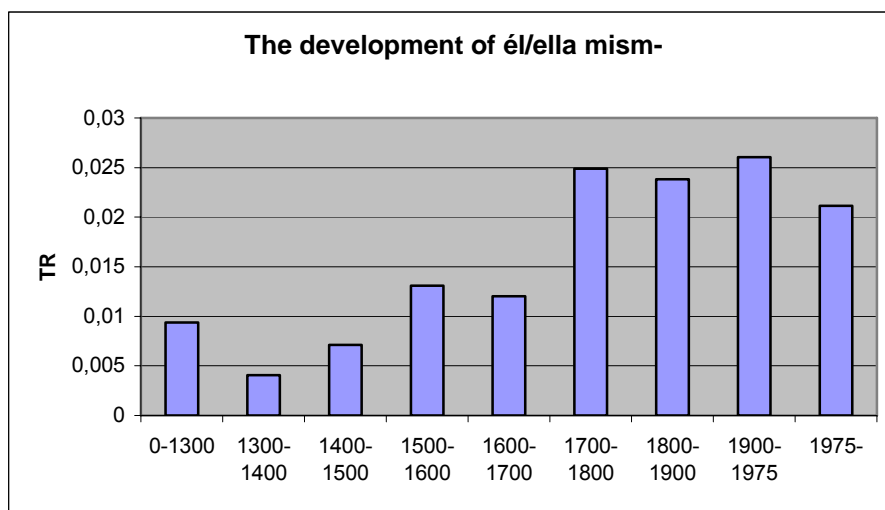


Figure 10 The development of *él/ella mism-* in terms of token ratio.

We can observe certain similarities by comparing the development of *él/ella (mism-)* with the patterns of change of the reflexive construction. In particular, the shift in token ratio from the fifteenth to the sixteenth century reminds us of the course of development of the reflexive construction. However, the change is much less drastic and in fact completely neutralized in the second part of the period, or in terms of the token ratio: the token ratio has been almost constant, and maybe even with a decreasing tendency, for the last four hundred years. It should also be noticed that the simple construction (*él/ella*), without intensification, is in Modern Spanish nearly 50 times more frequent than the intensified construction (the token ratio of *él/ella mism-* is 0,02). This tells us that the use of the basic intensifier-construction *él/ella mism-* has always been, and still is, relatively very lim-

ited. On the other hand, the use of the reflexive intensifier construction (the complex variant) is much more extended; see Table 1. The point I want to make here is that, according to the data presented in Table 14 and Figure 10, the dramatically increased use of the complex reflexive form cannot convincingly be ascribed to a general, pragmatically determined, usefulness of *mism-* as an intensifier of (pro)nominal phrases. At least, there is no indication that such a general tendency, driven by pragmatic principles of usefulness, should be the principal explanatory factor.

6.1.2 Increased speaker involvement

A paradigmatic interpretation of the data is, as we have seen, centered in the usefulness of the intensifier (*mism-*) as an optional device for foregrounding (pro)nominal phrases in discourse, cf. 2.1. Correspondingly, the use of the complex reflexive form involves, from a paradigmatic point of view, an optional promotion of the subject in discourse, and indirectly an increased speaker involvement. In Detges (2001), what he calls contrastive auto-thematization in spoken language is a related phenomenon, based on universal pragmatic and discourse principles (Detges 2001: 9; see also Detges 2002). The speaker marks that his contribution contrasts with a previous utterance, or with what the interlocutor presumably believes. This is shown in the following sequence, provided by Detges:

- (21) H1: [...] ¿Es imposible el entenderse entre dos generaciones diferentes, Adolfo Marsillach?
 H2: **Yo** [I] es que **creo** [I think] que no hace falta. Eh... Más bien creo que hace falta que no se entiendan. **Yo** [I] creo que está bien eso, porque la...hay un proceso...eh...competitivo, por el cual las generaciones nuevas necesitan afirmarse, y esa afirmación **yo creo** [I think] que se produce a través de un inmediato enfrentamiento con el enemigo más próximo, que evidentemente es el familiar que tienen delante. Pero **a mí** [to me] eso, insisto que no **me parece** [it seems] mal. (Corp. Oral 92, cit. Detges 2001: 9).

The sequence is taken from a television program in which the participants discuss the conflicts between the generations. The speaker is expressing various acts of contrastive auto-thematization in order to indicate that his view diverges from the general opinion. The interlocutor may, according to Detges, expect the content of this kind of utterances to be relevant in the Gricean sense since it is assumed to contain new information. This is a strong motive for using contrastive auto-

thematization as a rhetorical strategy; also in contexts with no contrasts of opinion at all, in which such a technique is not objectively justified. In the long run the abusive use of auto-thematization will lead to a loss of the original contrastive value in constructions that were meant as a way to indicate a high level of communicative relevance.

In spoken discourse, rhetorical strategies of auto-thematization may lead to the use of “heavy” focalizing constructions. Along with the emergence of the clitic reflexive marker in the sixteenth century, this may possibly have favored the use of *mism-* in the reflexive argument structure construction. More specifically, it could be argued that the loss of the original focalizing weight of the Spanish *a*-marker must have been a strong motivation for using another emphatic element, namely *mism-* (See also Section 2.2). Furthermore, we have seen that the optional intensifier *mism-* offers itself as a candidate for contrastive discourse promotion in the non-clitic construction as well, a communicative function that seems to be related to the auto-thematization, advocated by Detges (2001). See Section 2.1. In short, it might be hypothesized that discourse rhetorical strategies of this kind, spreading from first and second person to third person, will appear to be the principal explanatory force behind the increasing use of the complex form *sí mism-*. Given the validity of this hypothesis we may assume that the observed change, *ceteris paribus*, should be particularly extended in spoken language, as compared to written language. This assumption is based on the belief that speaker involvement, and the effect of intended pragmatic strategies, must tend to be particularly high in spoken discourse. In table 2, repeated for convenience, the data show, surprisingly, that there is not such an overall difference:

	Spoken language			Written language		
	sí	Sí mism-	TR	Sí	Sí mism-	TR
A = obj marker	3	102	34,00	170	4541	26,71
Para = to/for	16	11	0,69	1006	428	0,43
Por = by	149	80	0,54	2316	2104	0,91
De = of	51	49	0,96	1703	2950	1,73
En = in	271	105	0,39	1947	2447	1,26
Sobre = about	1	8	8,00	163	719	4,41
Con = with	79	16	0,20	2455	857	0,35

Table 2 Token ratio of *sí (mism-)* in spoken/written Modern Spanish

The prediction for the token ratio is almost fulfilled for the prototypical reflexive argument structure construction, principally *a sí mism-*, but also to some degree for *para sí mism-*. It should be taken into account, however, that the process of change favoring the use of the complex form *a sí mism-* is already completed at the beginning of the time period covered by the corpus CREA (1975), on which the comparison in Table 2 is exclusively based. Thus the data in Table 2, concerning *a sí mism-*, cannot be used directly to argue for the correctness of the hypothesis of usefulness. Since we do not have diachronic data of spoken Spanish, we will have to rely on the argument presented above and in 2.2: the loss of the focalizing weight of the Spanish *a*-marker could have favored the use of *mism-* in the reflexive argument structure construction.

In the more peripheral prepositional constructions the picture is markedly different. Only *sobre sí mism-* shows a relatively higher token ratio in spoken language. This specific case can probably be seen as a natural consequence of the type of meaning content that is usually involved in this construction.¹¹ Moreover, the token ratio of *sobre sí mism-* in spoken language is based on very few occurrences (9) in CREA. Regarding the rest of the prepositions, the token ratio for the complex form is strikingly lower in spoken language than in written language. Let me recall that this pattern would be markedly unexpected if rhetorical strategies, as described above, were the primary motive for the extended use of the complex construction.

The relatively sparse occurrences of the complex reflexive form in spoken language could alternatively be accounted for by applying simple principles of syntagmatic economy: speakers will tend to use shorter and/or simpler forms for more frequent contexts of use (Croft 2000: 142, Haiman 1983, Bybee 1985). From such principles we would, *ceteris paribus*, expect the simplest form, i.e. *sí*, to be the preferred one in spoken discourse, but not necessarily in written language.

It might be objected that since this study only concerns third person forms, and consequently only includes occurrences of *sí mism-*, while the first and second person forms have been excluded,

¹¹ The predicate frequently involves a kind of subjective reflection that makes the use of the *mism-* variant more compelling, as in the following example: La función del siquiatra no es la de "normalizar", sino tratar de dar al paciente una posible lucidez **sobre sí mismo**, una visión más clara sobre el mundo que le rodea [The role of the psychiatrist is not to "normalize" but to try to give the patient a possible lucidity upon himself, a clearer vision of the world that surrounds him] (CREA).

the outcome of written language occurrences has been biased to some extent. The reason for this bias in the data is that spoken language to a larger extent than written language involves first and second person forms. However, the third person forms of the unbound reflexive pronoun are certainly used in spoken language as well, and given the size of the modern corpus CREA (approx. 120 million words), enough material is provided for calculating a reliable token ratio for *sí mism-* in spoken language. But what is more important, the data shows that occurrences of first and second person complex forms, e.g. *para mí/ti mism-* (*to/for my/you self*) are surprisingly rare, even in spoken language, as compared to third person forms.¹² See Table 15:

	A	Para	Por	De	En	Sobre	Con
Mi	4485	1173	107	221	99	5	715
Ti	717	177	41	109	36	5	354
Si	3	16	149	51	271	1	79
mí mism-	38	5	8	9	5	2	18
ti mism-	15	2	9	2	4	0	4
sí mism-	102	11	80	49	105	8	16

Table 15 Complex forms in CREA, spoken language.¹³

It is therefore highly unlikely that the exclusion of first/second person forms should be responsible for the conclusion drawn on the basis of data in Table 2. We argued that the use of the complex form does not seem to be particularly extended in spoken language, as compared to written language, as predicted by the paradigmatic hypothesis. The impact of pragmatic strategies and speaker involvement, as described in this section, must be assumed to be particularly high in spoken language if this hypothesis were correct.

Another serious problem for the paradigmatic interpretation is that the hypothesis of increased speaker involvement, or auto-thematization, rests on universal discourse and pragmatic principles (Detges 2001: 9). It follows that if this interpretation were correct, we would expect to see the same

¹² This is in fact a general phenomenon. In languages that have reflexive pronouns in all three persons the use of the third person reflexive is much more frequent than the first or second person reflexive (Newmeyer 2003: 695).

¹³ The relatively high number of occurrences of the simple reflexive form in first and second person is due to the ambiguity of these forms. Searches for first and second person forms, e.g. *mí* or *ti*, also include non-reflexive constructions, for example: *no te preocupes por mí* [*don't worry about me*], or *Pedro ha comprado la tarta para ti* [*Pedro bought the cake for you*].

tendency, at least to some extent, in the German reflexive intensifier-construction. However, there is no evidence that a similar tendency should have been observed in German (see e.g. König & Siemund 2000: 63f, Baker 1995: 98f, Hole 2002: 133ff; Eckardt 2001, 2002, Siemund 2000).

A convincing analysis of the development of the Spanish reflexive should preferably account for the fact that the development in Spanish differs from the development in German since the German reflexive-intensifier distinction is typologically very similar to the one we find in the Spanish non-clitic construction. The difference between the two languages must be due to language specific factors. Generalization patterns in Spanish, for instance, may have an influence on the Spanish construction that is stronger than the impact of some universal-typological pragmatic principles. In this perspective, the development of the construction in Spanish may be analyzed by drawing on specific conditions for Spanish, which have no parallel in German, such as the prototypical clitic reflexive construction and its close interaction with the non-clitic construction. See Section 6.2.

All together this leads us to conclude that the data show only little evidence for the claim that the increased use of the complex reflexive form in non-clitic constructions should be interpreted in terms of pragmatically determined usefulness. The pragmatic impact on the development should therefore not be exaggerated. This conclusion applies not only for the quantitative data for the prototypical intensifier-construction (6.1.1), but also for the comparison of written and spoken language (6.1.2), and the typological prediction concerning German. The influence from the clitic argument structure construction, in which the complexity of the unbound form has become compulsory, seems to be relevant for the evolution of the more peripheral non-clitic construction.

6.2 *A syntagmatic explanation: reanalysis*

We are certainly not claiming that the observed change in usage may not involve an increased use of reflexive intensification due to pragmatic factors. However, the analysis in Section 6.1 suggests that pragmatic motivations for the increased use of the complex construction cannot per se explain the data. We will now examine the possibility that the observed change in the use of *mism-* involves not merely an entrenchment of the optional reflexive intensification, but a true reanalysis of the

construction (*sí* + intensifier *mism-* being reinterpreted as a complex reflexive *sí mism-*) in the mental grammar of the user. Notice that in French the negative *pas* was originally part of an intensifier construction *ne (pas)*, much like Modern English *not a bit*. Later it was reanalyzed as a part of a complex negative *ne pas* (see e.g. Croft 2000: 130f; Labov 1994: 596f; Jespersen 1917: 7f). Similar changes are attested in many languages (Campbell 1998: 240). In this Section it will be hypothesized that the intensifier construction *sí (mism-)* is in a process of being reanalyzed in a similar way.

Recall that concerning the clitic argument structure construction, in which the clitic *se* (=reflexive) has become the principal co-referential device, we have hypothesized that *a sí (mism-)* has been reanalyzed as a reflexive intensifier *a sí mism*, see Section 2.2. Here the expectation is that an increasing token ratio of the clitic construction *se...a sí...* is correlated with an increasing token ratio of the complex form *a sí mism-*.¹⁴ For both the clitic and non-clitic construction the hypothesis of reanalysis implies that in the mental grammar of the user, *mism-* is being conceived as obligatory in the complex form *sí mism-*.

The question is now: what would be a plausible motivation for reanalyzing *sí (mism-)* in the non-clitic construction as a complex reflexive *sí mism-* if it cannot be exclusively a matter of pragmatic usefulness (cf. the analysis and discussion in section 6.1)? Tentatively we may argue that the preference of the complex form in the prototypical clitic construction (*se... a sí mism-*) could simply have spread to non-clitic construction types by means of syntactic generalization. If this is true, we must expect the increasing token ratio of the complex form *a sí mism-* to be correlated with a general increase in the token ratio for the construction type ‘preposition + *sí mism-*’.

In sum, there are two expectations:

Expectation A: the increasing token ratio of the clitic construction (*se*) is correlated with an increasing token ratio of the complex form *a sí mism-*.

Expectation B: the increasing token ratio of the complex form *a sí mism-* is correlated with a general increase in the token ratio for the construction type ‘preposition + *sí mism-*’.

¹⁴ The token ratio of the clitic construction *se...a sí...* reflects the clitic use as related to the non-clitic use, and the token ratio of the complex form *a sí mism-*, as defined in Section 3, the use of the complex form *a sí mism-* as related to the simple form *a sí*.

In Table 16 the token ratio for the development of the clitic construction, contrasted with the token ratio for *a sí mism-*, is tabulated¹⁵, cf. expectation A. In order to focus graphically on the early development of the clitic construction, calculations of the inverse token ratio: *a sí / se...a sí* (instead of *se...a sí / a sí*) and *a sí / a sí mism-* (instead of *a sí mism- / a sí*) are preferable. Notice the shift from the 15th to the 16th century: 12,74 → 3,52 and 8,58 → 2,06 respectively:

(Se)...a sí (mism-)	Inverse TR for <i>se...a sí</i> ...	Inverse TR for <i>a sí mism-</i>
0–1300	31,00	1,21
1300–1400	33,00	3,86
1400–1500	12,74	8,58
1500–1600	3,52	2,06
1600–1700	1,65	1,20
1700–1800	1,16	0,56
1800–1900	1,09	0,43
1900–1975	1,00	0,14
1975–	0,93	0,04

Table 16 Inverse token ratio for *se...a sí* and *a sí mism-*

In Figure 11 this development is represented graphically. The data show that the development of the clitic construction and the complex *a sí mism-* construction, in terms of token ratio, are strongly correlated confirming the expectation A:

¹⁵ The calculation is based on registered occurrences of the word *se* within a distance of five words from the expression *a sí*.

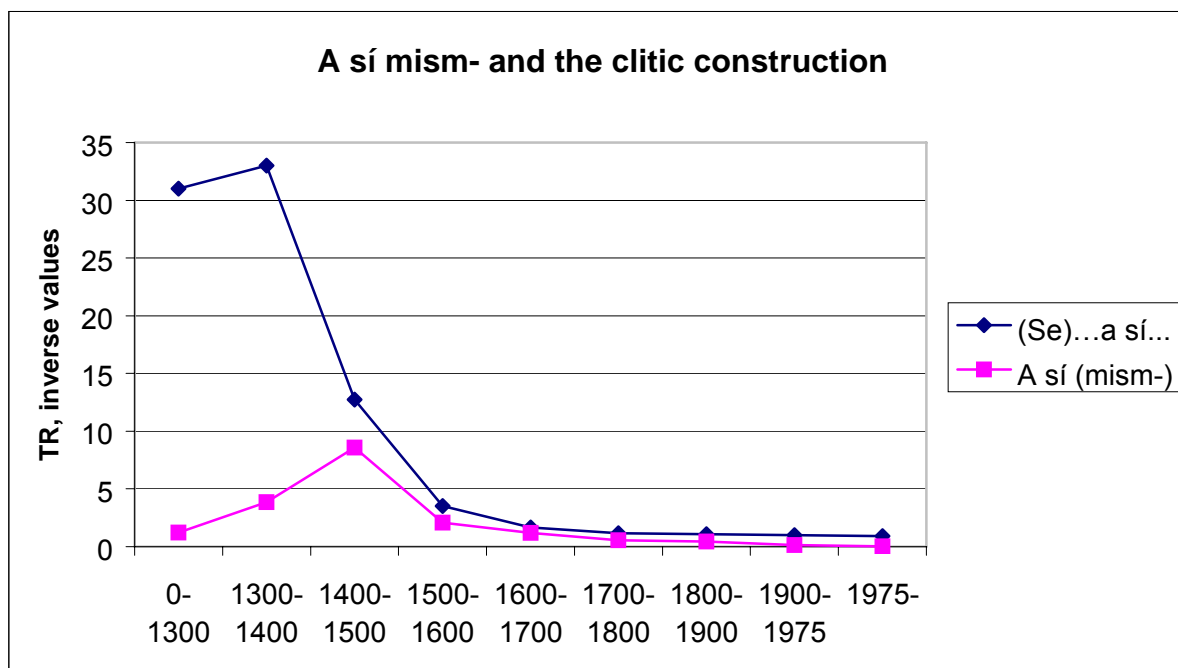


Figure 11 The development of the complex *a sí mism-* as related to the clitic construction.

As regarding expectation B, when comparing the development of *a sí mism-* with the prepositional construction in general, we found that the observed changes in the use of the various prepositional non-clitic constructions follow quite systematically the development of the *a sí mism-* construction. The result of this comparative analysis, tabulated in Table 17 and represented in Figure 12, confirms B:

Prep.+ <i>sí</i> + <i>mism-</i>	Inv TR for <i>a sí mism-</i>	Inv TR for 'prep. + <i>sí mism-</i> '
0–1300	1,21	25,75
1300–1400	3,86	19,64
1400–1500	8,58	9,45
1500–1600	2,06	6,22
1600–1700	1,20	5,69
1700–1800	0,56	5,25
1800–1900	0,43	2,55
1900–1975	0,14	1,36
1975–	0,04	1,05

Table 17 Inverse token ratios for *a sí mism-* and prep. + *sí mism-*

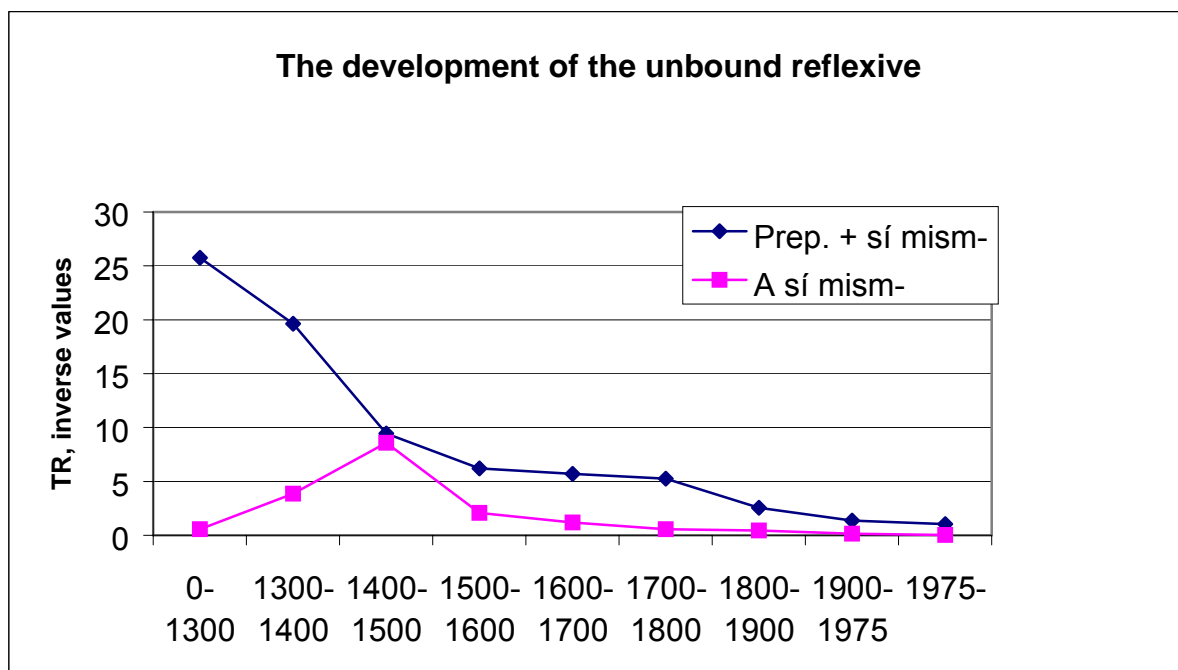


Figure 12 The development of the unbound reflexive

In Table 18 the paradigmatic and the syntagmatic hypothesis have been compared by calculating correlations in the change of token ratio between ‘preposition + *sí* + *mism-*’ and ‘*el/ella mism-*’, ‘*a sí mism-*’ respectively, cf. 6.1.1:

Preposition + <i>sí</i> (<i>mism-</i>):	De	En	Para	Por	Con	Sobre
Correlation with <i>El/ella</i> (<i>mism-</i>)	0,589982	0,609848	0,539366	0,771493	0,656156	0,403208
Correlation with <i>a sí</i> (<i>mism-</i>)	0,922158	0,902376	0,89185	0,779116	0,848604	0,995476

Table 18 Correlations of the change of token ratio. Correlation quotients.

The changes in token ratio for ‘preposition + *sí mism-*’ fit much better with those for ‘*a sí mism-*’ (the correlation quotients are generally closer to 1) than with those for ‘*él/ella mism-*’. Hence, the calculation of correlation quotients confirms expectation B.

To summarize the argument: In the syntagmatic perspective the observed change has been interpreted as a manifestation in usage of a reanalyzed reflexive expression. Due to the cliticisation of *se*, the reflexive expression *a sí (mism-)* has been reanalyzed as a complex intensifier *a sí mism-* in

the argument structure construction. This syntactic pattern is generalized by spreading to non-clitic, non-prototypical construction types, motivating the reanalysis of ‘preposition + *sí* + optional *mism*’ as a complex reflexive expression: ‘preposition + *sí mism*-’. The entire chain of changes can be schematized as follows:

Change 1: Cliticization of *se* \Rightarrow

Change 2: Reanalysis: *a sí (mism-)* \rightarrow *se (a sí mism-)* in clitic constructions \Rightarrow

Change 3: Syntactic generalization: *sí (mism-)* \rightarrow *sí mism-* \Rightarrow

Change 4: Reanalysis: *prep. sí (mism-)* \rightarrow *prep. sí mism-* in non-clitic constructions

The syntagmatic interpretation is supported by general observations on language change made by William Labov (1994). He observed that in actual speech one variant is rarely chosen over another for functional reasons, that is, as a result of an intentional individual choice. Instead he argues that the choice is much more mechanical, and syntagmatic in nature, forming part of what he calls *systemic readjustment*, without regard to communicative needs. As one example he talks about a tendency to preserve parallel structures in successive sentences, a mechanism which is similar to the syntactic generalization proposed as change 3 (ibid.: 550). Labov does certainly not say that functional arguments are never valid. Rather, he claims that the pressure from specific communicative needs is relatively weak, and can be overridden by a variety of other factors. See also the discussion in Newmeyer 2003.

This analysis relies tacitly on the assumption that a syntactic generalization (change 3) may trigger the loss of a semantic distinction and a pragmatic option (change 4). From a mainstream functional (usage based) point of view the analysis might therefore appear problematic, though it certainly raises some interesting questions: Is a clear cut semantic distinction between reflexivization and intensification in Modern Spanish maintainable? Is it possible for syntactic changes to trigger semantic changes?

Carl L. Baker (1995) has shown that the distinction between locally binding properties, i.e. reflexivization, and properties of discourse prominence, i.e. intensification, tends to be blurred to the user. If this is so, the clear cut semantic distinction between reflexivization and intensification is not maintainable, and a reanalysis of reflexive intensification, *sí (mism-)*, as a complex reflexive, *sí mism-*, is semantically and pragmatically highly motivated. There seems, thus, in the syntagmatic perspective to be a semantic-pragmatic motivation as well as a syntactic motivation for the reflexive marker + intensifier to be reanalyzed. Furthermore, the process of reanalysis will naturally tend to eliminate the explicitly coded meaning of intensification; cf. the understanding of reanalysis adopted in this paper:

Form-function reanalysis is syntagmatic: it arises from the (re)mapping of form-function relations of combinations of syntactic units and semantic components. The process may nevertheless have an apparently paradigmatic result, for example, a change of meaning of a syntactic unit. (Croft 2000: 120)

I am not suggesting that syntactic changes in general trigger semantic changes, but simply that in this specific case a syntactic generalization seems to have had a significant impact on the process of change. Given the validity of Baker's argumentation, the question whether a similar semantic-pragmatic motivation for reanalysis might have had an impact on the German reflexive as well becomes relevant. There is, apparently, no evidence, as we have pointed out, that a similar process of change is taking place in German. This suggests that the hypothesized reanalysis of the Spanish non-clitic reflexive may be strongly influenced by the syntactic change in the clitic construction, as proposed in this paper, since the clitic construction is specific for Spanish as opposed to German. On the other hand, this study should be replicated for other Romance languages with clitic reflexives, for instance Italian, to see if the same patterns can be observed. If this is the case, it will provide further support for the analysis offered in this paper.

7 Conclusion

The data show a continuous increase in the use of *mism-* in Spanish reflexive constructions during the last 700 years. The change from the fifteenth to the sixteenth century has been shown to be particularly striking. Assuming that the approach must be usage based, two different interpretations of the data have been discussed:

A) The paradigmatic interpretation, leading to the hypothesis:

The complex construction *sí mism-* is becoming more entrenched in the grammar as a useful device for intensification.

B) The syntagmatic interpretation (of reanalysis), leading to the hypothesis:

Mism- is becoming part of the reflexive marker.

B in particular has shown to fit well with empirical data on the development of the reflexive construction, though A-phenomena, to some extent, may have been motivational factors for reanalysis (B); cf. Section 2.2 and 6.1. On the other hand, we have argued that the hypothesized reanalysis of the non-clitic construction seems to be influenced by the syntactic structure of the prototypical reflexive argument structure construction (the clitic construction). If this interpretation of the data is correct, a complex reflexive marker, *sí mism-*, is in a process of being grammaticalized in Spanish, which has not hitherto been realized in Spanish linguistics.

References

- Alarcos Llorach, E. 1995. *Gramática de la Lengua Española*. Real Academia Española. Colección Nebrija y Bello. Madrid: Espasa Calpe.
- Baker, C. 1995. "Contrast, discourse prominence, and intensification, with special reference to locally free reflexives in British English". *Language* 71(2). 63-101.
- Bello, A & R. J. Cuervo. 1954. *Gramática de la lengua Castellana*. Prepared by N. Alcalá-Zamora y Torres. Buenos Aires Sopena Argentina.

- Bosque, I. & Violeta Demonte, eds. 1999. *Gramática Descriptiva de la Lengua Española (GDE)*. Madrid: Espasa.
- Burzio, L. 1991. "The morphological basis of anaphora". *Journal of Linguistics* 27. 81-105.
- Burzio, L. 1996. "The role of the antecedent in anaphoric relations". *Current Issues in Comparative Grammar*. Dordrecht: Kluwer. 1-45.
- Bybee, J. 1985. *Morphology: a study into the relation between meaning and form*. Amsterdam: John Benjamins.
- Bybee, J., R. D. Perkins & W. Pagliuca. 1994. *The evolution of grammar: tense, aspect and modality in the languages of the world*. Chicago: University of Chicago Press.
- Bybee, J. & P. Hopper, eds. 2001. *Frequency and the emergence of linguistic structure*. Amsterdam: John Benjamins.
- Campbell, L. 1998. *Historical Linguistics*. Edinburgh: Edinburgh University Press.
- CORDE. Internet-based database: <http://www.rae.es/>. The Royal Spanish Academy (RAE).
- Corominas, J. 1954-57. *Diccionario crítico etimológico de la lengua castellana*. Madrid: Gredos.
- CREA. Internet-based database: <http://www.rae.es/>. The Royal Spanish Academy (RAE).
- Croft, W. 2000. *Explaining Language Change. An Evolutionary Approach*. London: Longman.
- Delille, K. H. 1970. *Die geschichtliche Entwicklung des präpositionalen Akkusativs im Portugiesischen*. Bonn: Romanisches Seminar.
- Detges, U. 2001. In press. "La gramaticalización de los acusativos preposicionales en las lenguas iberorrománicas. Una hipótesis pragmática". *Variación sintáctica en español: un reto para las teorías de la sintaxis*, ed. by V. Bellosta von Kolbe & G. Knauer. Tübingen: Niemeyer.
- Detges, U. 2002. "Du sujet parlant au sujet grammatical. L'obligatorisation des pronoms sujets en ancien français dans une perspective pragmatique". Submitted paper".
- Eckardt, R. 2001. "Reanalyzing selbst". Manuscript. To appear in *Natural Language Semantics*.
- Eckardt, R. 2002. "Semantic change in grammaticalization". G. Katz et al 2002. 53-67.
- Everaert, M. 1986. *The syntax of Reflexivization*. Dordrecht: Foris.
- Fernández Ramírez, S. 1987. *Gramática Española, 3.2. El pronombre* (1951). Vol. Preparado por

- José Polo (2. edición.). Madrid: Arco/Libros.
- Frajzyngier, Z. & T. S. Curl, eds. 2000. *Reflexives. Forms and functions*. Amsterdam/Philadelphia: John Benjamins.
- Franc, J. A. & J. M. Blecula. 1975. *Gramática Española*. Barcelona: Ariel.
- Geniusiene, E. 1987. *The typology of reflexives*. Berlin: Mouton de Gruyter.
- Haiman, J. 1983. "Iconic and economic motivation". *Language* 59, 781-819.
- Haiman, J. 1985. *Natural syntax*. Cambridge: Cambridge University Press.
- Hole, D. 2002. Agentive Selbst in German. G. Katz et al. 2002. 133-150.
- Huang, Y. 2000. *Anaphora. A Cross-linguistic Approach*. Oxford: Oxford University Press.
- Jespersen, O. 1917. *Negation in English and other languages*. Copenhagen: A. F. Høst.
- Katz, G, S. Reinhard & P. Reuter, eds. 2002. *Sinn & Bedeutung VI, Proceedings of the Sixth Annual Meeting of the Gesellschaft für Semantik*. University of Osnabrück.
- Keller, R. 1994. *On language change: the invisible hand in language*. London: Routledge.
- Kemmer, S. 1995. "Emphatic and reflexive –self: expectations, viewpoint, and subjectivity". *Subjectivity and subjectification* ed. by Dieter Stein & Susan Wright, 55-82. Cambridge: Cambridge University Press.
- König, E. 1991. *The meaning of Focus Particles: A comparative perspective*. London: Routledge.
- König, E. & P. Siemund. 1996a. „Emphatische reflexiva und fokusstruktur“. *Sprache und Pragmatik* 40 ed. by Inger Rosengren. Lund: Lunds Universitet. 1-42.
- König, E. & P. Siemund. 1996b. „Selbst-Reflektionen“. *Wenn die Semantik Arbeitet. Festschrift für Klaus Baumgärtner* ed. by Gisela Harras. Tübingen: Niemeyer. 277-302.
- König, E. & P. Siemund. 2000. "Intensifiers and reflexives: A typological perspective". Frajzyngier & Curl 2000. 41-74.
- König, E. 2001. "Intensifiers and Reflexive Pronouns". *Language Typology and Language Universals: An international Handbook* ed. by M. Haspelmath, E. König, W. Oesterreicher & W. Raible. Berlin: Walter de Gruyter.
- Labov, W. 1994. *Principles of linguistic change, vol. 1: Internal factors*. Oxford: Blackwell.

- Lehmann, C. 1985. *Grammaticalization: synchronic variation and diachronic change*. *Lingua e Stile* 20. 303-18.
- Lightfoot, D.W. 1991. *How to set parameters: arguments from language change*. Cambridge: Cambridge University Press.
- Meier, H. 1945. "O problema do acusativo preposicional no catalão". *Boletim de Filologia* VIII. 327-60.
- Milroy, J. 1992. *Linguistic variation and change*. Oxford: Blackwell.
- Newmeyer, F. J. 2003. "Grammar is grammar and usage is usage". *Language* 79: 4, 682-707.
- Otero, C. P. 1999. "Pronombres reflexivos y recíprocos". Ignacio Bosque & Violeta Demonte. 1999. 1427-1517.
- Pedersen, J. (forthcoming). "Reflexive intensification in Spanish. Toward a complex reflexive?". *Selected Papers from the XVIth International Conference on Historical Linguistics*, Copenhagen 2003, ed. by Michael Fortescue, Eva Skafte Jensen, Jens Erik Mogensen & Lene Schøsler. John Benjamins: Amsterdam.
- Pica, P. 1991. "On the interaction between Antecedent-Government and Binding: The case of Long-distance Reflexivization". *Long-distance Anaphora* ed. by Jan Koster & Eric Reuland. Cambridge: CUP. 119-36.
- Quirk, R., S. Greenbaum, G. Leech & J. Svartvik. 1985. *A comprehensive Grammar of the English Language*. London: Longman.
- Real Academia Española. 1981. *Esbozo de una Nueva Gramática de la Lengua Española*. Madrid: Espasa-Calpe, S.A.
- Reinhard, T. & E. Reuland. 1993. "Reflexivity". *Linguistic Inquiry* 24(4). 657-720.
- Schladt, M. 2000. "The typology and grammaticalization of reflexives". Frajzyngier & Curl 2000. 103-124.
- Siemund, P. 2000. *Intensifiers. A Comparison of German and English*. London: Routledge.
- Torrego, E. 1995. "From Argumental to Non-Argumental Pronouns: Spanish Doubled Reflexives". *Probus* 7. 221-241.

Van Hoek, K. 1997. *Anaphora and conceptual structure*. The University of Chicago Press.

Zribi-Hertz, A. 1995. “Emphatic or Reflexive”? On the Endophoric Character of French lui-même and Similar Complex Pronouns”. *Journal of Linguistics* 31. 333-374.